Domain-Guided Novelty Detection for Autonomous Exploration

David R. Thompson

Jet Propulsion Laboratory, California Institute of Technology

4800 Oak Grove Dr. Pasadena, CA

david.r.thompson@jpl.nasa.gov

Abstract

In this work, novelty detection identifies salient image features to guide autonomous robotic exploration. There is little advance knowledge of the features in the scene or the proportion that should count as outliers. A new algorithm addresses this ambiguity by modeling novel data in advance and characterizing regular data at run time. Detection thresholds adapt dynamically to reduce misclassification risk while accommodating homogeneous and heterogeneous scenes. Experiments demonstrate the technique on a representative set of navigation images from the Mars Exploration Rover "Opportunity." An efficient image analysis procedure filters each image using the integral transform. Pixel-level features are aggregated into covariance descriptors that represent larger regions. Finally, a distance metric derived from generalized eigenvalues permits novelty detection with kernel density estimation. Results suggest that exploiting training examples of novel data can improve performance in this domain.

1 Introduction

Novelty detection aims to identify subsets of data that are inconsistent with the majority [Markou and Singh, 2003; Hodge and Austin, 2004; Grimaldi *et al.*, 2006]. Standard approaches include estimating data's probability density [Quinn and Williams, 2007; Markou and Singh, 2003], characterizing its geometry, or identifying its support [Scholkopf *et al.*, 2001; Chen *et al.*, 2008]. Conventionally there are three different versions of novelty detection that each exploit different kinds of prior information [Hodge and Austin, 2004]. Often prior examples of "novel" and "regular" data are available, and novelty detection reduces to a two-class classification task [Hodge and Austin, 2004]. If only examples of regular observations are present then novelty detection becomes a one-class classification problem. Here one identifies observations that differ significantly from the training set [Chen *et al.*, 2008]. Finally, novelty detection with no training examples is tantamount to unsupervised classification or density estimation. One identifies outlier observations that are far from the
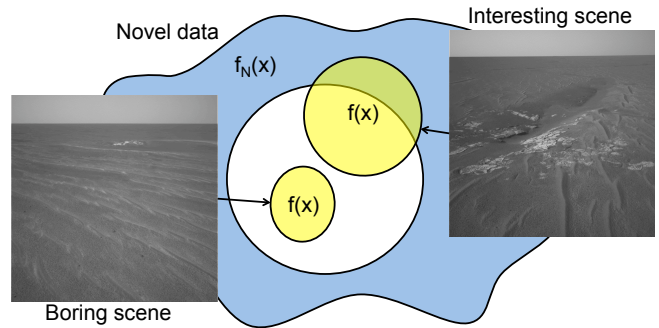

Figure 1: Density estimation for novelty detection in Mars Exploration Rover Images. Each image patch is associated with a feature vector x. Terrain varies significantly across scenes and a different proportion of each image counts as "novel." The proposed solution adapts the novelty detection threshold to minimize expected misclassification risk based on a common, shared model $f_N(x)$ of novel objects. Images courtesy NASA, Cornell, Caltech Jet Propulsion Laboratory.

majority [Chen *et al.*, 2008], or likely to come from a different distribution [Quinn and Williams, 2007].

This work examines a fourth kind of novelty detection that models novel observations in advance and characterizes regular observations at run time. This is appropriate for applications that do not simply seek outliers for their own sake, but instead use some other standard to define which data points count as novel. In this case there is some presumed cost for false positive detections, and the proportion of novel observations could vary across trials. An advance model of novel observations can leverage domain knowledge, adjusting the novelty detection decision appropriately for each new trial.

Our motivating application is autonomous exploration by remote agents such as orbiting spacecraft, submersibles, or planetary rovers. These robots must explore unknown environments autonomously while only rarely communicating their discoveries to operators. Here novelty detection can identify interesting image features for prioritized transmission [Castaño *et al.*, 2003]. If only a handful of data products can be transmitted due to bandwidth constraints, novelty detection can maximize the number of interesting terrain features in the downlink. It can suggest scientifically inter-

esting image areas to inform region-of-interest based image compression [Wagstaff *et al.*, 2004]. Novelty detection can also identify promising features for automatic deployable instruments [Castaño *et al.*, 2007; Smith *et al.*, 2007]. Often sensors like spectrometers or high resolution cameras are expensive in time, bandwidth or energy; novelty detection can suggest targets for these instruments and improve the overall science return of autonomous operations.

In remote exploration one cannot always anticipate the proportion of observations that should count as novel, i.e. that will be of interest to the operators. Consider the two images from a Mars Rover traverse [Maki *et al.*, 2003] in Figure 1. Here the goal is to target novel features that are unlike others in the image. The first "boring scene" is a featureless dunescape where even the most extreme outliers make poor science targets. The second image exemplifies the opposite extreme, where a significant portion of the image counts as interesting. For instance, the rock outcrops and textured sediment could all justify followup observations. An appropriate novelty detection rule must adapt across scenes in order to detect the appropriate portion of outlier image features.

The proposed solution leverages prior models of novel observations to minimize expected misclassification loss. An expert provides an advance library of example targets; this suggests appropriate novelty detection decision rules for each new environment the agent explores. This work demonstrates the method with a simple novelty detection algorithm based on density estimation. The agent sets detection thresholds dynamically to reduce misclassification risk. These detection decisions adapt to homogeneous and heterogeneous scenes.

Exploration agents' limited computational resources preclude the most intensive image analysis techniques. Planetary exploration is particularly constrained since it relies on low power, radiation certified flight processors whose progress lags years or decades behind consumer devices [Cox, 2007]. We address these constraints with efficient image analysis using the integral image transform [Lienhart and Maydt, 2002]. A convolution with Haar-wavelet like filters [Lienhart and Maydt, 2002] produces local image features. We aggregate local features into covariance descriptors that represent larger image patches, and detect novel image patches with a distance metric based on generalized eigenvalues [Tuzel *et al.*, 2006].

This paper evaluates novelty detection techniques on a representative set of navigation images taken by the Mars Exploration Rover "Opportunity" during its multiple-year traverse of the Meridiani Plateau, Mars [Maki *et al.*, 2003]. Here we detect novel rock and sediment features in rover surface images. Tests investigate the performance of adaptive thresholding, and the image descriptors in general, for both simulated target selection and prioritized return tasks. They demonstrate improved performance using prior domain knowledge to guide novelty detection decisions.

2 Adaptive Thresholding Approach

The proposed method treats novelty detection as a binary classification task, assigning a cost to each false positive (a regular datapoint labeled as novel) and false negative (a missed novel datapoint). This leads naturally to a Bayesian approach in which the utility of a decision rule is based on its expected classification loss. Given a new dataset one can choose an appropriate novelty threshold to minimize the expected loss of the detection decision.

Consider the data $X = \{X_1, X_2, \ldots X_n\}, X_i \in \mathcal{X}$ drawn from a two-component mixture model. The first component corresponds to the majority or regular data; it appears with probability P_R and has PDF $f_R(x)$. A novel component appears with probability $P_N = 1 - P_R$ and has PDF $f_N(x)$. The PDF of the joint dataset $f(x)$ decomposes:

$$f(x) = P_R f_R(x) + P_N f_N(x) \qquad (1)$$

A binary decision rule $D_\theta(x) : \mathcal{X} \mapsto [0, 1]$ classifies unlabeled features. It evaluates to 1 if the feature is novel and 0 otherwise. For example, the tests that follow use a decision rule based on the PDF; novel data points are those with low probability density. Alternatives include distance- or geometry-based measures of novelty. One can parameterize the decision rule by some θ capturing choices like the detection threshold or the basis used to represent data points. The classification loss L_θ is a linear combination of false positives and false negatives weighted by coefficients L_α and L_β respectively. The expected loss is:

$$
\begin{aligned}
E[L_\theta] & = L_\alpha P_R \int D_\theta(x) f_R(x) dx + \\
& \quad L_\beta P_N \int (1 - D_\theta(x)) f_N(x) dx \qquad (2)
\end{aligned}
$$

The best choice for θ is that which minimizes this expression. Dropping constant terms produces an objective function $\mathcal{V}_\theta$:

$$
\begin{aligned}
\mathcal{V}_\theta & = L_\alpha P_R \int D_\theta(x) f_R(x) dx - \\
& \quad L_\beta P_N \int D_\theta(x) f_N(x) dx \qquad (3)
\end{aligned}
$$

Domain knowledge can estimate terms of this equation that are shared across trials. For example, in the one class classification version of novelty detection one presumes that the distribution of regular data is constant and models $f_R(x)$ entirely in advance. This is not appropriate for the current application since scene structure and illumination conditions could vary significantly across images. Instead, one can more safely assume that *novel* data points are drawn from some broad shared distribution. For reducing risk it is sufficient to estimate $f(x)$, which is evident from runtime observations, and $f_N(x)$, which is common across all environments in the application domain.

Substitution from equation 1 removes references to $f_R(x)$; and nonparametric plug-in estimates [Wasserman, 2006] replace the remaining PDFs. The density $\hat{f}(x)$ consists of point masses at the set of new observations, and $\hat{f}_N(x)$ of point masses given by the training set of novel data.

$$
\begin{aligned}
\mathcal{V}_\theta & = L_\alpha \int D_\theta(x) \hat{f}(x) dx \\
& \quad -(L_\alpha + L_\beta) P_N \int D_\theta(x) \hat{f}_N(x) dx \qquad (4)
\end{aligned}
$$

This expresses the objective in terms of S_θ, the fraction of observations detected as novel in the new test data. Similarly $S_{N\theta}$ is the fraction of novel training points that would successfully be detected as novel in context of the new test data.

$$\mathcal{V}_\theta \;\; \propto \;\; L_\alpha\, S_\theta - (L_\alpha + L_\beta)P_N S_{N\theta} \qquad (5)$$

This motivates the adaptive thresholding procedure (Algorithm 1). It requires a training library N of novel examples, the learned or assumed proportion P_N of novel data in the test environment, and penalties L_α and L_β for false positive and false negative detections. Then one chooses an optimal detection threshold $\theta^\star$ to minimize equation 5. Small changes to θ that do not alter any classifications do not affect this objective, so it is never necessary to evaluate a number of thresholds greater than the total size of the training and test sets.

Input: Training set of novel data N, test set X
      Prior probability P_N of novel data,
      False positive loss L_α
      False negative loss L_β
Output: Detected "novel" subset of X

foreach *possible threshold θ* **do**
   $S_\theta = 0, S_{N\theta} = 0$;
   foreach *x in X* **do**
      if $D_\theta(x) = 1$ **then**
         $S_\theta = S_\theta + \frac{1}{|X|}$;
   foreach *x in N* **do**
      if $D_\theta(x) = 1$ **then**
         $S_{N\theta} = S_{N\theta} + \frac{1}{|N|}$;
$\theta^\star = \operatorname{argmin}_\theta [L_\alpha S_\theta - (L_\alpha + L_\beta)P_N S_{N\theta}]$;
return *x for which $D_{\theta^\star}^\star(x) = 1$*

Algorithm 1: Novelty detection with adaptive thresholds.

Figure 2 shows a simple example based on a mixture of two Gaussians. Regular data is drawn from a narrow distribution $\mathcal{N}(\mu, \sigma^2), \mu \sim \mathcal{N}(0,1), \sigma \sim \mathrm{gamma}(1,1)$. The novel data comes from a broad Gaussian $\mathcal{N}(\mu_N, \sigma_N^2)$ with $\mu_N \sim \mathcal{N}(0,1)$ and $\sigma_N \sim \mathrm{gamma}(9, 0.5)$. Here the decision rule fits a single Gaussian distribution to $f(x)$ by maximum likelihood, and classifies as novel those points that fall under a density threshold.

This simulation weights false positives and false negatives equally, so $L_\alpha = 1.0$. The threshold θ is set presuming $P_N = 0.05$, when in fact it varies over the range shown. Loss as a function of P_N is shown for two novelty detection strategies. The first decision rule uses a single static threshold of $\theta = 0.05$, while the second uses the adaptive threshold of Algorithm 1. Figure 2 shows the average results of over 100 trials. The adaptive threshold outperforms conventional novelty detection, incurring a lower mean loss per trial over a wide range of probabilities including the case where the prior is correct ($P_N = 0.05$).

Figure 3 illustrates a similar test where P_N is held constant but the loss function varies. Adaptive thresholding improves performance, especially in regimes where false positive penalties are weighted very differently from false negatives. The algorithm can compensate by returning more or fewer detections as appropriate.
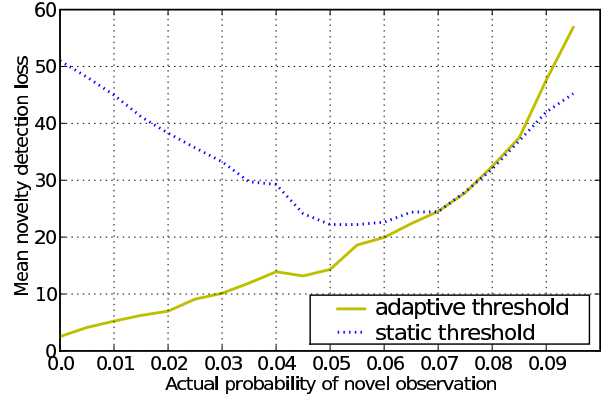

Figure 2: Novelty detection with adaptive thresholding. Performance profiles describe static and adaptive decision rules that presume $P_N = 0.05$, when the actual probability varies. Empirical loss is given by equation 5 with $L_\alpha = L_\beta = 1$. The mean of 100 trials is shown.
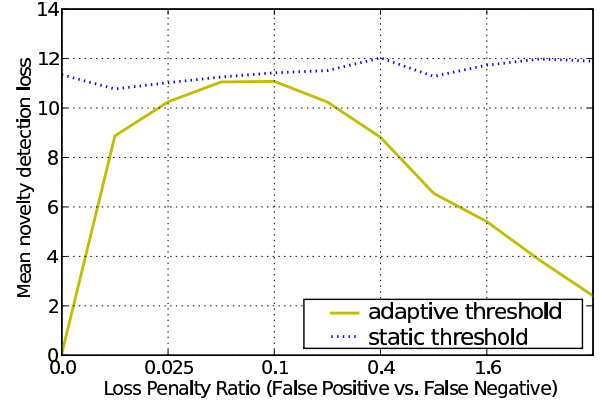

Figure 3: Novelty detection with adaptive thresholding. Performance is shown for static and adaptive decision rules when $P_N = 0.05$, and loss functions have varying ratios for L_α and L_β. Lines show empirical loss (mean over 100 trials).

3 Image Analysis

Figure 4 shows the basic architecture for analyzing Mars Exploration Rover images. The procedure divides the 1024×1024-pixel image into a grid of 50×50 pixel subwindows. Here our automatic instrument targeting requires classifying each subwindow as "ordinary" or "novel" with respect to the other subwindows in the image. The original Navigation Camera image (Top Left) contains many irrelevant image elements such as the horizon, farfield features, and the rover deck. Range masking uses stereo data to discard these features; it excludes all subwindows whose pixels lie outside the range from $2.5m$ to $10m$. Limiting analysis to these ranges also prevents distance-induced changes in focus and texture.

The next step computes a set of "Haar wavelet-like" filters that correspond roughly to oriented edge, bar, boxcar and center-surround filters [Lienhart and Maydt, 2002]. These re-
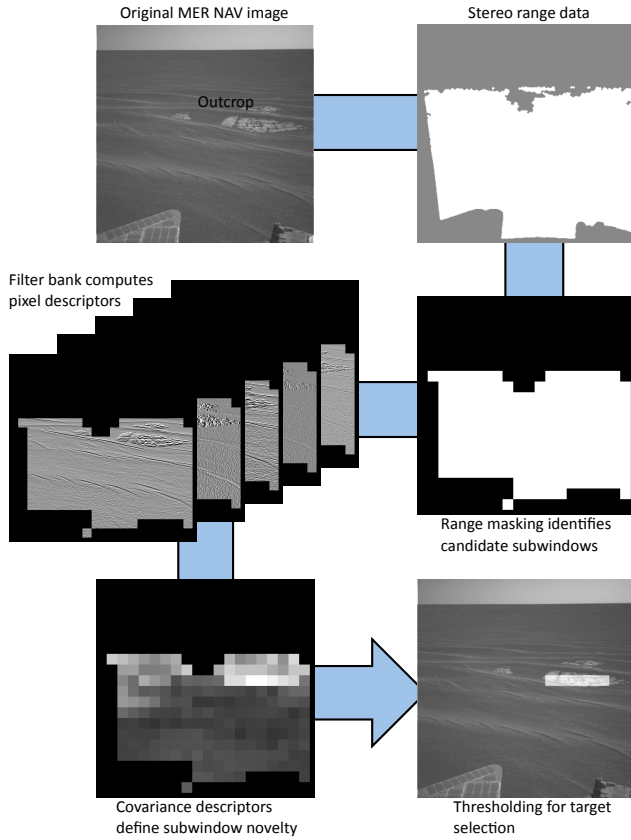
Figure 4: Image analysis and novelty detection architecture. The original image (Top Left) is convolved with a filter bank to produce pixel-level features. Range data (Top Right) excludes irrelevant parts of the image such as the extreme far field and the rover deck. Pixel-level features combine into a single covariance descriptor for each subwindow (Bottom Left). Adaptive thresholding identifies novel subwindows.

veal pixel properties like albedo, the local first and second image derivatives, edges and corners. The integral image transform permits fast calculation of these values using a handful of integer operations. It also furnishes an efficient contrast normalization detailed by Lienhart et al [2002]. The result for each pixel is a numerical feature vector with one dimension for each filter channel. These experiments use five Haar-wavelet-like filters: a "center-surround" blob detector, and vertical and horizontal gradient and edge filters. All Haar-wavelet-like filters have a width of 5 pixels. This feature set was determined manually through trial and error. Automatic feature selection could also be used to explore a more complete filter bank parameterization.

Detecting outlier subwindows requires a numerical representation for the content of these larger image regions. Image texture has been found in other contexts to be a good correlate of geomorphology in rover images [Thompson *et al.*, 2008]. Here we characterize subwindows with covariance descriptors, discussed at length in Tuzel et al [2006]. The covariance descriptor represents a rectangular region by the covariance

matrix of its pixel-level features. It is a compact yet discriminative representation of texture with some invariance to the spatial arrangement of features in the subwindow. It is especially advantageous for remote operations due to its computational efficiency. The integral image transform can quickly compute the required summations to produce covariance matrix entries from pixel features.

We use generalized eigenvalues to compare multiple descriptors [Förstner and Moonen, 1999]. For covariance matrices C_1 and C_2, the generalized eigenvalues $\{\lambda_i(C_1, C_2)\}_{i=1...m}$ are the solutions to the expression $\lambda_i C_1 v_i - C_2 v_i = 0$, for associated eigenvectors v_i. Following Tuzel et al, a suitable distance metric is:

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^{m} \ln^2 \lambda_i(C_1, C_2)} \qquad (6)$$

This distance metric makes the covariance descriptors relevant for a wide range of novelty detection strategies including distance-based approaches or support-based strategies using Mercer kernels.

Pairwise distances permit density estimation with a squared exponential kernel [Wasserman, 2006]. For a set of n subwindows, the probability density for a particular subwindow x is the following:

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sqrt{2\pi}w} \exp\left\{ -\frac{\rho(C_x, C_j)^2}{2w^2} \right\} \qquad (7)$$

The parameter w is a bandwidth that regularizes the density estimate. It can be set through methods such as the Normal reference rule or Bayesian approaches. These experiments favor leave-one-out cross validation [Wasserman, 2006].

The resulting novelty detection function estimates $\hat{f}(x)$ for the subwindows of each new image and classifies those with values below some threshold as novel. We identify the best threshold by computing the density estimate $\hat{f}(x)$ for each training example in the context of the new image. This leads to risk values (equation 5) for thresholds that exclude varying numbers of subwindows. The best threshold balances conservatism (reducing the loss due to false positive detections) and recall (detecting as many of the training subwindows as possible). Thus, adaptive thresholding produces few or no detections in instances where the image is quite homogeneous relative to the training data, the prior probability of a novel point is low, or false positives detections have a high cost.

4 Case Study: MER Navigation Images

Here a case study tests the proposed approach using images from the Navcam instrument [Maki *et al.*, 2003] onboard the Mars Exploration Rover (MER) "Opportunity." This rover has traveled several kilometers across the Meridiani plateau in a mission lasting over five years. It has crossed terrain consisting of dunes punctuated by occasional rock outcrops and other novel sediment features. The case study considers 11 representative images from diverse locations along the traverse, excluding images with overlapping content, extreme lighting or very poor contrast. Filtering for range yields a set of over 400 subwindows.
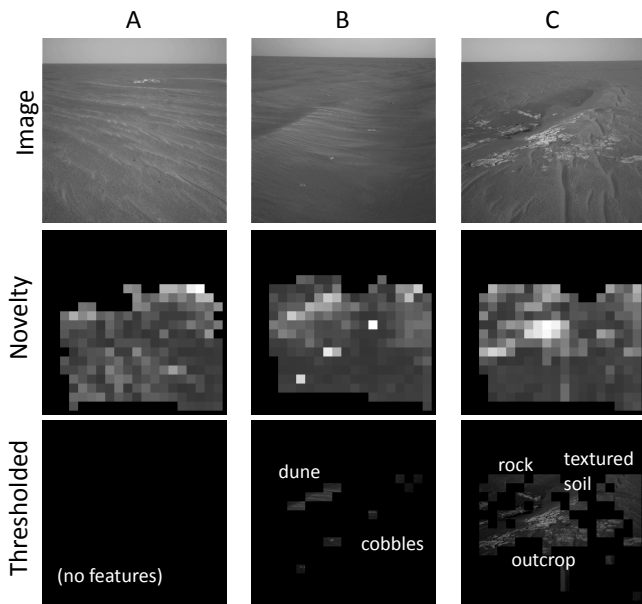
Figure 5: Typical novelty detection results demonstrating the adaptive density threshold. A: Nondescript dunes; no novel subwindows are detected. B: Several small rocks are present. The algorithm identifies these, along with the crest of a dune, as novel. 14 subwindows fall under the density threshold. C: A large portion of this area is covered by soil and outcrop features similar to the novel examples from the training set. The resulting threshold detects over 100 subwindows.

In practice mission scientists themselves may disagree on objectives [Castaño *et al.*, 2005], so one cannot claim with certainty that a detected feature is *not* interesting. However, we can get some idea of performance with specific objects that are unquestionably novel and easy to identify. Many scenes in the data set contain a small portion of outcrop or surface rocks; for the purposes of these trials, novel subwindows are those containing obvious rocks or rock outcrops greater than 20 pixels in length. These subwindows comprise a training set of novel data and give a lower bounds on the number of true positives.

Figure 5 shows typical results with the adaptive density threshold. The second row illustrates the density $\hat{f}(x)$ computed for each subwindow using kernel density estimation. Brighter values correspond to higher novelty, or equivalently, low probability density. The final row shows subwindows selected by adaptive thresholding. The scene in column A is relatively homogeneous and the adaptive threshold does not detect any image subwindow as novel. In column B several small rocks are present. The algorithm identifies these as novel, along with the crest of a dune. 14 subwindows fall under the density threshold. Finally, a large portion of the column C image is covered by soil and outcrop features. The resulting threshold detects over 100 subwindows.

Figure 6 quantifies performance for the entire dataset. It compares detection results to a static thresholding scheme that applies the same threshold across all images. This test
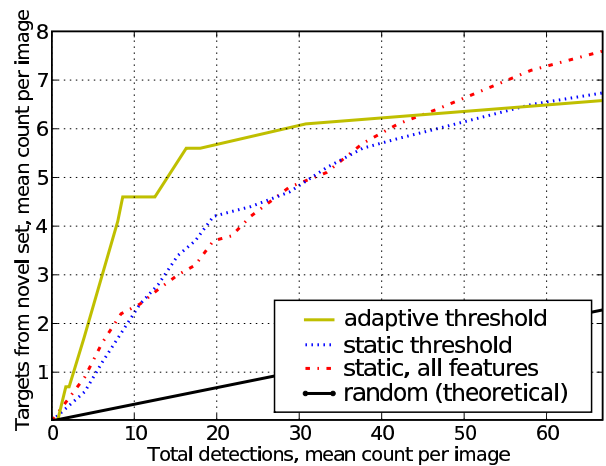

Figure 6: Test results on MER imagery for both static and adaptive thresholds using the center-surround feature. A third line shows novelty detection using the entire feature set with a fixed threshold. Finally, the solid dark line illustrates the theoretical performance of a selective sampling and return strategy that does not consider image content. Vertical and horizontal axes correspond to the number of detected novel subwindows that appear in the set of rock outcrops.

uses a truncated feature set comprised solely of the center-surround filter (a basic edge detector) from the Haar-like filter bank. This is important because the training set is relatively sparse, and the complete feature set provides very discriminative covariance features leaving little overlap between training and test distributions. Then the estimate of $\hat{f}_N$ is poor, and trivially low thresholds can exclude all novel subwindows in the training set. Limiting analysis to the single most informative filter remedies this problem.

Figure 6 suggests that, with respect to detecting rock features, both novelty detection methods offer significant advantages over naïve instrument targeting that ignores image content. Adaptive thresholding successfully returns many more outcrop subwindows than the static method for a wide range of loss functions. It offers nearly double the detection precision of the static method when few detections are returned. The third (red) line of Figure 6 shows the results of the novelty detection method using the entire feature set and a fixed detection threshold. This offers little benefit over the center-surround filter, but there may be some improvement in the extreme high recall regimes. Future work with a larger dataset may investigate this phenomenon for adaptive thresholding.

5 Conclusions

This work demonstrates a nonparametric strategy for optimizing novelty detection decision rules based on a prior model of novel data. A case study applies the technique to adaptive thresholding for image-based novelty detection in the context of remote science. The integral image furnishes efficient feature descriptors for rectangular image regions. Then, density estimation based on a generalized eigenvalue metric identifies novel regions. Simulations and field data suggest adap-

tive thresholding can improve novelty detection performance for a range of environments and loss functions.

There is considerable scope for future research. This preliminary work does not consider any novelty detection strategies beyond density estimation. Additionally, it may be useful to evaluate alternative parameterizations besides the decision threshold. One can adapt the feature vector itself, searching with each new dataset for the representation that best discriminates between regular and novel observations. Kernelized novelty detection methods might tackle this problem by parameterizing the kernel function directly.

The previous section alluded to practical difficulties of estimating $f_N(x)$. In particular, a highly discriminative feature set can reduce overlap between the training and test domains, making it trivially easy for a very low threshold to exclude the entire training set of novel subwindows. This follows naturally from the assumption that $\hat{f}_N$ accurately represents the novel data, which is increasingly difficult to enforce for finite training sets as the feature space becomes more discriminative. As a consequence performance is sensitive to noninformative dimensions of the feature vector. Fortunately optimal density estimation from thin data is tantamount to a regularization problem with classic remedies such as Bayesian priors and cross validation.

Acknowledgments

References

[Castaño et al., 2003] R. Castaño, M. Judd, RC Anderson, and T. Estlin. Machine learning challenges in Mars rover traverse science. In *ICML Workshop on Machine Learning Technologies for Autonomous Space*, 2003.

[Castaño et al., 2005] R. Castaño, K. Wagstaff, L. Song, and RC Anderson. Validating rover image prioritizations. *The Interplanetary Network Progress Report*, 42-160, 2005.

[Castaño et al., 2007] R. Castaño, T. Estlin, R.C. Anderson, D.M. Gaines, A. Castaño, B. Bornstein, C. Chouinard, and M. Judd. OASIS: Onboard autonomous science investigation system for opportunistic rover science. *Journal of Field Robotics*, 24(5), 2007.

[Chen et al., 2008] Y. Chen, X. Dang, H. Peng, and H. Bart. Outlier Detection with the Kernelized Spatial Depth Function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18, 2008.

[Cox, 2007] Brian Cox. Updated avionics electronic trends. *Jet Propulsion Laboratory, Interoffice Memorandum (Internal Distribution)*, OPFMS:033, 2007.

[Förstner and Moonen, 1999] W. Förstner and B. Moonen. A Metric for Covariance Matrices. *Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University*, 1999.

[Grimaldi et al., 2006] M. Grimaldi, B. Espey, and P. Cunningham. Mining the SDSS Database. *Trinity College Dublin, Computer Science Report No. TCD-CS-2006-01, Trinity College Dublin.*, 2006.

[Hodge and Austin, 2004] Victoria J. Hodge and Jim Austin. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22:85126, 2004.

[Lienhart and Maydt, 2002] Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. *International Conference on Image Processing*, 2002.

[Maki et al., 2003] JN Maki, JF Bell III, KE Herkenhoff, SW Squyres, A. Kiely, M. Klimesh, M. Schwochert, T. Litwin, R. Willson, A. Johnson, et al. Mars Exploration Rover Engineering Cameras. *Journal of Geophysical Research*, 108(E12):12–1, 2003.

[Markou and Singh, 2003] M. Markou and S. Singh. Novelty Detection: A Review (2 Parts). *Signal Processing*, 83(12):2481–2497, 2003.

[Quinn and Williams, 2007] J.A. Quinn and C.K.I. Williams. Known Unknowns: Novelty Detection in Condition Monitoring. *Lecture Notes in Computer Science*, 4477:1, 2007.

[Scholkopf et al., 2001] B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[Smith et al., 2007] T. Smith, D. R. Thompson, D. Wettergreen, N. Cabrol, K. Warren-Rhodes, and S. Weinstein. Life in the Atacama: Science Autonomy for Improved Data Quality. *Journal of Geophysics Research*, 112,G04S03, Dec 2007.

[Thompson et al., 2008] D. R. Thompson, T. Smith, and D. Wettergreen. Information-Optimal Selective Image Return for Autonomous Rover Traverse Science and Survey. In *International Conference on Robotics and Automation*, 2008.

[Tuzel et al., 2006] O. Tuzel, F. Porikli, and P. Meer. Region Covariance: A Fast Descriptor for Detection and Classification. In *European Conference on Computer Vision*. Springer, 2006.

[Wagstaff et al., 2004] K. I. Wagstaff, R. Castaño, S. Dolinar, M. Klimesh, and R. Mukai. Science-Based Region-of-Interest Image Compression. *35th Lunar and Planetary Science Conference, Abstract 1934*, 2004.

[Wasserman, 2006] Larry Wasserman. *All of Nonparametric Statistics*. Springer, 2006.